

Exercício em sala Ori TF*IDF

Nome: Maycon Douglas Batista dos Santos

Matricula: 11921BSI209

Questão a)

termo	D1 (f)	D2 (f)	D3 (f)	D4 (f)	D5 (f)
logan	1	1	0	1	0
ororo	1	0	1	0	0
stark	0	1	1	0	0
parker	0	2	0	0	0
groot	0	0	1	3	2
rocket	0	0	1	0	2
x-men	1	0	0	1	0
vingadores	0	1	2	0	0
guardiões	0	0	2	0	1
Total tf	3	5	8	5	5

termo	D1 f	D2 f	D3 f	D4 f	D5 f	D1-tf	D2-tf	D3-tf	D4-tf	D5-tf
logan	1	1	0	1	0	1	1	0	1	0
ororo	1	0	1	0	0	1	0	1	0	0
stark	0	1	1	0	0	0	1	1	0	0
parker	0	2	0	0	0	0	2	0	0	0
groot	0	0	1	3	2	0	0	1	2.58	2
rocket	0	0	1	0	2	0	0	1	0	2
x-men	1	0	0	1	0	1	0	0	1	0
vingadores	0	1	2	0	0	0	1	2	0	0
guardiões	0	0	2	0	1	0	0	2	0	1
Total tf	3	5	8	5	5	-	-	-	-	-

termo	D1-tf	D2-tf	D3-tf	D4-tf	D5-tf	IDF	Tf*idf	Tf*idf	Tf*idf	Tf*idf	Tf*idf
logan	1	1	0	1	0	0,74	0,74	0,74	0	0,74	0
ororo	1	0	1	0	0	1,32	1,32	0	1,32	0	0
stark	0	1	1	0	0	1,32	0	1,32	1,32	0	0
parker	0	2	0	0	0	2,32	0	4,64	0	0	0
groot	0	0	1	2.58	2	0,74	0	0	0,74	1,91	1,48

rocket	0	0	1	0	2	1,32	0	0	1,32	0	2,64
x-men	1	0	0	1	0	1,32	1,32	0	0	1,32	0
vingadores	0	1	2	0	0	1,32	0	1,32	2,64	0	0
guardiões	0	0	2	0	1	1,32	0	0	2,64	0	1,32
Total tf	-	-	-	-	-						

Questão B)

D1 = (ororo, x-mex)

D2 = (parker)

D3 = (vingadores, guardiões)

D4 = (groot)

D5 = (Rocket)

R: são os termos mais relevantes já que não são tão frequentes nos documentos então no que eles aparecem recebem um peso maior para a busca.

Questão C)

D1 = (logan)

D2 = (logan)

D3 = (groot)

D4 = (logan)

D5 = (guardiões)

R: São termos que são muitos frequentes dentre todos os documentos assim eles recebem uma relevância maior.

Questão D)

Será um valor 0, por exemplo tenho 10 documentos e o termo 'mesa' aparece em todos então seria $\log_2(10/10) = 0$ como o tf-idf é uma multiplicação independente do valor de tf como o idf será 0 então o resultado final será 0.